

Predicting a place to open a Clothing store in Toronto neighbourhood.

Kiruthika Thirunavukkarasu

August 2020

1. Introduction:

A person is trying to open a clothing store in neighborhood of Toronto and is looking in for some places where it would be good to open. Before opening a clothing store, it is better to know about the places where there is a demand for such store. It is also good to know about other clothing stores which are already present in nearby places and about their ratings. When we identify the places which does not have much of clothing stores, it would be good to choose that place.

2. Data Requirements:

- The Data required will be collected from Wikipedia.
- https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M_,
- The data consists of postal codes and the respective places.
- Some cleaning process will be done in order to make the data clear and usable.
 - *Ignore cells with a borough that is Not assigned
 - *Combining rows which has more than one neighborhood one postal code area
 - *Getting the latitude and Longitude values and assigning it to a new columns in Data frame.
- Using Foursquare API, the neighborhoods and venues in the neighborhoods will be discovered. Then the number of clothing store in each venue will be taken into account.
- Using Clustering algorithms, the clusters will be formed in the neighborhoods.
- Finally, the folium library is used to visualize the clusters. Based on the clusters formed, prediction can be done as where the demand is high to open a clothing store.

3. Methodology:

In our case, the problem is to find some places in the neighborhood of Toronto where it will be best to open clothing store. The venues which are already present in neighborhoods are first listed. The number of venues are also categorized. In all the neighborhoods, the total number of venues and most common venues in the respective venues are identified. Using Hierarchical clustering, the venues in neighborhoods and distance among them are visualized.

Using Kclusters and agglomerative clustering, the neighborhoods are partitioned as 6 clusters and most common venues in those clusters are identified. We can visualize the same using map. Finally, the venues present in the 6 clusters are identified and in our case, clothing store is not present in any of the clusters. Hence it will be good to open the clothing store in any of the six clusters in the neighborhoods.

4. Result:

From the methodologies we used to explore the neighborhoods of Toronto, we are able to identify that compared to other venues, clothing stores are less in number and hence can be opened in many locations since there is more demand.

5. Discussion:

All the five clusters suffers from Clothing store. Hence it is best to open in any of the five clusters of the neighbourhood.

6. Conclusion:

Details of places in Toronto like postal code, Neighbourhood, latitude, Longitude we were able to visualize the map of Toronto. Using Foursquare the venues which are present in the neighbourhoods and their ratings are also identified. Based on the results, we were able to find out the most common venues and their occurrence in each neighbourhood. From these details, we can also visualize the venues in map. Using these details as reference it is easy to predict a place where it will be good to open a clothing store. Similarly, to open any such venues same procedure can be followed and predicted.